PAII-NLP at SMM4H 2021: Joint Extraction and Normalization of Adverse Drug Effect Mentions in Tweets

Zongcheng Ji, Tian Xia and **Mei Han**
PAII Inc.
3000 El Camino Real, Palo Alto, CA 94306
{jizongcheng, SummerRainET2008, hanmei613}@gmail.com

Abstract

This paper describes our system developed for the subtask 1c of the sixth Social Media Mining for Health Applications (SMM4H) shared task in 2021. The aim of the subtask is to recognize the adverse drug effect (ADE) mentions from tweets and normalize the identified mentions to their mapping MedDRA preferred term IDs. Our system is based on a neural transition-based joint model, which is to perform the recognition and normalization simultaneously. Our final two submissions outperform the average F1 by 1-2%.

1 Introduction

With the popularity of social media such as Twitter, people often publish messages online in regard to their health such as the information related to the adverse drug effects (ADEs). Mining such type of information from social media is helpful for pharmacological post-marketing surveillance and monitoring. The aim of the sixth Social Media Mining for Health Applications (SMM4H) shared task in 2021 (Magge et al., 2021) is to mining such invaluable health information from social media. We participate in the subtask 1c of SMM4H 2021, which is to recognize the ADE mentions from tweets and normalize the identified mentions to their mapping MedDRA [1] preferred term IDs.

2 Task and Data Description

We give the formal definition of the end-to-end task. Briefly, given a tweet x published by a user, and a knowledge base (KB, i.e., MedDRA) which consists of a set of concepts, the goal of the task is to identify all the ADE mentions $M = \{m_1, m_2, ..., m_{|M|}\}$ in x and to link each of the identified mention m_i to the mapping MedDRA preferred term ID e_i in KB, $m_i \rightarrow e_i$. If there is no mapping concept in KB for m_i, then $m_i \rightarrow NIL$, where NIL denotes that m_i is unlinkable.

Table 1: Overall statistics of the dataset.

	#tweets	#mentions	#unique concepts
trn	17,375	1,706	317
dev	915	86	57
tst	10,984	-	-

Table 1 shows the statistics of the dataset provided by the organizers. We use the training (trn) and development (dev) sets to build our system and submit the predictions on the testing (tst) set.

We use MedDRA v21.1 as the KB, which consists 25,463 unique preferred term IDs.

3 The Approach

Preprocessing. We preprocess all the tweets with the following steps (Ji et al., 2016): 1) tokenize the tweets with whitespace and punctuations; 2) lowercase the tokens; 3) replace the urls with "httpurl"; 4) replace the @user with "username"; 5) replace the escape characters with their original form (e.g., & $\rightarrow$ &).

We preprocess all the mentions and concepts in KB with the following steps (Ji et al., 2020): 1) replace the numerical words to their corresponding Arabic numerals (e.g., one / first / i / single $\rightarrow$ 1); 2) tokenize the mentions and concepts with whitespace and punctuations; 3) remove the punctuations; 4) lowercase the tokens.

Neural Transition-based Joint Model. We cast the end-to-end task as a sequence labeling task and convert the whole task as an action sequence prediction task. We follow previous studies of applying Neural Transition-based Model for named entity recognition (NER) (Lample et al., 2016; Wang et al., 2018) with SHIFT, OUT, REDUCE, SEGMENT actions for the recognition purpose and further extend the model by adding LINKING actions for the normalization purpose.

[1] https://www.meddra.org/

Proceedings of the Sixth Social Media Mining for Health Workshop 2021, pages 126–127
June 10, 2021.

Table 2: Results on the development set.

	Precision	Recall	F1
Submission 1	0.623	0.545	0.582
Submission 2	0.570	0.557	0.563

Table 3: Results on the test set.

	Precision	Recall	F1
Submission 1	0.331	0.179	0.230
Submission 2	0.317	0.196	0.240
Average	0.231	0.218	0.220

Input Representation We represent each token x_i in a tweet x by concatenating its character-level word representation, non-contextual word representation, and contextual word representation:

$$x_i = [v_i^{char}; v_i^w; ELMo_i] \tag{1}$$

where v_i^{char} denotes its character-level word representation learned by using a CNN network (Ma and Hovy, 2016), v_i^w denotes its non-contextual word representation initialized with Glove (Pennington et al., 2014) embeddings, which is pre-trained on a large-scale Twitter corpus of two billion tweets, and $ELMo_i$ denotes its contextual word representation initialized with ELMo (Peters et al., 2018).

Search and Training For efficient decoding, a widely-used greedy search algorithm (Lample et al., 2016; Wang et al., 2018) is adopted to minimize the negative log-likelihood of the local action classifier, *i.e.,* to minimize the cross-entropy loss between the output distribution with the gold-standard distribution:

$$\mathcal{L}(\theta) = -\sum_t log\ p(a_t|r_t) \tag{2}$$

where θ denotes all the parameters in this model.

4 Results and Conclusions

We submit the following two results with two different strategies:

- **Submission 1**: single model result with the neural transition-based joint model.
- **Submission 2**: voting result with 5 best single model results.

We report the Precision, Recall and F1 for each ADE extracted where the spans overlap either entirely or partially AND each span is normalized to the correct MedDRA preferred term ID.

Table 2 and 3 show the evaluation results on the development and test sets, respectively. Average denotes the arithmetic median of all submissions made by all the teams participate the end-to-end subtask. Results show that the proposed method outperform the average F1 by 1-2%.

In the future, we will further tune the model and explore other popular contextual word representations learned from BERT (Devlin et al., 2018).

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Zongcheng Ji, Aixin Sun, Gao Cong, and Jialong Han. 2016. Joint Recognition and Linking of Fine-Grained Locations from Tweets. In *WWW*, pages 1271–1281.

Zongcheng Ji, Qiang Wei, and Hua Xu. 2020. BERT-based Ranking for Biomedical Entity Normalization. In *AMIA 2020 Informatics Summit*, pages 269–277.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural Architectures for Named Entity Recognition. In *NAACL*, pages 260–270, San Diego, California. Association for Computational Linguistics.

Xuezhe Ma and Eduard Hovy. 2016. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. In *ACL*, pages 1064–1074, Berlin, Germany. Association for Computational Linguistics.

Arjun Magge, Ari Klein, Ivan Flores, Ilseyar Alimova, Mohammed Ali Al-garadi, Antonio Miranda-Escalada, Zulfat Miftahutdinov, Eulàlia Farré-Maduell, Salvador Lima López, Juan M Banda, Karen O'Connor, Abeed Sarker, Elena Tutubalina, Martin Krallinger, Davy Weissenbacher, and Graciela Gonzalez-Hernandez. 2021. Overview of the Sixth Social Media Mining for Health Applications (# SMM4H) Shared Tasks at NAACL 2021. In *Proceedings of the Sixth Social Media Mining for Health Applications Workshop & Shared Task*.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *EMNLP*, pages 1532–1543.

Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep Contextualized Word Representations. In *NAACL-HLT*, pages 2227–2237.

Bailin Wang, Wei Lu, Yu Wang, and Hongxia Jin. 2018. A Neural Transition-based Model for Nested Mention Recognition. In *EMNLP*, pages 1011–1017. Association for Computational Linguistics.